JPL D-17849

August 5, 1999

Rev. 0

A Recommendation for a JPL Core Meta-Data Specification

Knowledge Management Standards Working Group

J. A. U'Ren, Working Group Leader

Jet Propulsion Laboratory
California Institute of Technology

CONTENTS

Executive Summary

The Problem

With recent breakthroughs in computing technology and associated standards, we have witnessed a rapid increase in the amount and variety of on-line information available via the Internet and within organizational intranets. The down-side of this rapid growth is an information glut in which finding relevant information is often difficult and unwieldy. Network resources such as "web-crawlers" and "spiders" (commonly referred to as Internet search engines) can index the full-text of HTML pages. These resources have proven useful, but returns from Internet search engines often contain too much information, return irrelevant information, miss relevant information that is not Web-friendly, and don't communicate the pedigree of the information. Also, search engines can be "spoofed" or "spammed," i.e. tricked, through a variety of techniques. Using an Internet search engine is often like trying to get a drink from a fire-hose – it's more than you can handle, it's difficult to actually get what you want, and you really don't know where the hose has been and whether you should be drinking water from it.

Mechanisms that structure and associate diverse types of on-line information could enable search engines to provide more timely and accurate results. Standardized meta-data can provide the structure that will enable a number of capabilities in terms of publishing, indexing, searching, retrieving, and archiving information.

What is Meta-data?

Meta-data is data about data. It describes the attributes and contents of an original document or work. Real life examples of meta-data are a TV guide (meta-data) that provides information about TV programs (real data) or a library card catalogue (meta-data) that provides information about books in a library (real data). When used effectively, well structured meta-data makes information accessible by labeling its contents consistently, leaving a pathway for users to find the information they need.

Standardizing and embedding meta-data in real data enables the information producers to associate the meta-data related to the information they are producing. In the past meta-data entry has been left to a keypunching or data input group. Often this has led to disconnections and time lags between the publishing of real data and its associated meta-data. In today's world of rich object formats, the producers of information can include relevant meta-data embedded (or encapsulated) in the information as it is published.

Embedded meta-data can be automatically or semi-automatically updated every time the real data is updated. Meta-data can be embedded in many data types commonly used, e.g., MS-Office formats such as Word DOC, Excel XLS, PowerPoint PPT as properties; HTML files as MetaTags, Adobe PDF files as properties, GIF and PNG graphic files in Description fields, STEP Part 21 engineering data files in the header section, etc.

Establishing Core Meta-data

The first step in implementing a standardized meta-data is to establish a set of meta-data elements common to all data used and published in a community. This common set of elements is referred to as *core meta-data*. Developing a common meta-data set will be a starting point for all data which can then be extended for specific data sets as required.

After reviewing a number of meta-data specifications, the Knowledge Management Standards Working Group recommends the Dublin Core (DC) Element Set for use at JPL. Each of the meta-data specifications reviewed have viable strengths and useful applications, but a key factor in recommending a core meta-data specification is to select one that is not narrowly focused and could accommodate a community of diverse data types as represented at JPL.

> **Note:** selecting the Dublin Core Meta-data Specification does not preclude the use of any other meta-data naming schemas to extend DC as a discipline or an application requires. The DC Element Set is simply a starting point from which the JPL community can begin. Many current implementations of DC include extensions for specific applications in specific disciplines.

The DC Element Set is a simple set of 15 descriptive elements that can be used to describe network resources (e.g. web pages, images, multimedia objects, engineering models, etc.) .

A growing acceptance of the Dublin Core Element Set is due to:

- Simplicity - very simple to learn
- Flexibility - every element is repeatable and optional
- Extensibility - can be refined and enhanced for more complex applications
- Integration - can be embedded invisibly in HTML and MS Office documents
- Support - recognized by the World Wide Web Consortium with broad international support.

> **Note:** the Dublin Core Element Set includes a feature that can tag information about access restrictions. The Rights attribute is designated to communicate copyright restrictions, terms of access, conditions for reuse, etc. ***JPL could use this attribute to flag information that is ITAR sensitive.***

Associating DC meta-data structure with all information produced is easily accomplished. It is possible for individuals and projects to begin including DC meta-data using tools that are already in use. Commonly used tools such as the Microsoft Office Suite (Word, Excel, PowerPoint, etc.) or most other MicroSoft Windows95/NT applications can embed "custom properties" that can contain the exact DC meta-data labels and attributes, and these custom properties can be part of document templates that are made available to all users within the JPL community.

An important implementation issue that deserves consideration is the development of a JPL schema that defines how the Dublin Core Element Set can be extended for use by all disciplines at JPL. Considerable work is being done to demonstrate the extensibility of the DC element set by various commercial, government, and academic organizations.

Recommendation

Structured and standardized meta-data is one part of an overall architectural philosophy that enables producers and users of information at JPL. It can simplify the ability to input information into repositories; it can enable better indexing of published information; it can provide improved searching capabilities; it can improve retrieval of information stored in repositories; and it can enhance the reuse of information.

This recommendation for the adoption of a Lab-wide Core Meta-data specification will not create a perfect world of information with perfectly formatted meta-data. Adoption of the DC Element Set is another step toward enabling information created at JPL to be interconnected and related in significantly improved ways. Using this approach, we can create a significantly better, a richer and more structured environment in which information that adheres to standard conventions allow many automated functions and leverage many new capabilities that flow from this structured environment.

1. Introduction

1.1 Purpose

The purpose of this document is to provide a preliminary recommendation for a JPL core meta-data specification. This document is not a specification - - it recommends the adoption of a specification. This document is not an implementation plan - - it recommends the adoption of a core meta-data specification that has been implemented at numerous sites around the world and can be implemented at JPL.

1.2 Scope

The scope of this document is to outline a core meta-data specification that can be used in administrative, engineering and science domains at JPL.

1.3 Goals

The goals of a common, JPL-wide core meta-data specification are to provide common attributes for administrative, engineering and science information created and published that will improve:

- Indexing
- Searching
- Retrieval
- Referencing
- Access
- Input
- Extraction
- Reuse

and to include mechanisms that provide extensions unique to categories of information in administrative, engineering and science disciplines.

2. Meta-Data Modeling Concepts

Well-structured meta-data makes information accessible by labeling its contents consistently. Meta-data describes the attributes and contents of an original document or work. Put simply, meta-data is data about data.

A more detailed definition is:

> *"data that describes other data; any file or database that holds information about another database's structure, attributes, processing or changes."*

> (**Note:** although the term is widely used without the dash, the word *Metadata* is a registered trademark of Metadata Information Partners).

The World Wide Web Consortium (W3C) asserts that meta-data has two main functions: (1) to provide a means to discover that the data set exists and how it might be obtained or accessed and (2) to document the content, quality, and features of a data set and so give an indication of its fitness for use.

Tim Berners-Lee, recognized as the "father" of the Web, defines meta-data as "machine understandable information about web resources and other things". He goes on to say:

> "The phrase 'machine understandable' is key. We are talking about information which software agents can use in order to make life easier for us, ensure we obey our principles, the law, check that we can trust what we are doing, and make everything work more smoothly and rapidly. In the future, when meta-data languages and engines are more developed, it should also form a strong basis for a web of machine understandable information about anything: about people, things, concepts and ideas. We keep this fact in our minds in the design, even though the first step is to make a system for information about information."

In developing standardized meta-data methodologies, the W3C and other standards groups have created distinctions through the use of the following terms:

- Resource – an identifiable object of interest that is accessible and available to the public
- Meta-data – a set of properties or attributes that aggregates some description of a resource
- Binding – the mechanism that makes the association between meta-data and the resource it describes.

Meta-data and the resource it describes may be separate entities (e.g. a library catalogue card and the book it describes) or embedded in the resource (e.g. information contained in a book's title and copyright page about the book). In the first case the association is implicit but in the later case there must be a mechanism for associating the meta-data and the real data.

A binding makes the association between meta-data and the resource it describes. In the library analogy, the binding may be a Dewey-number that points a person from the catalogue to the locations of the book on a bookshelf. On the Internet, a binding occurs via a URL that identifies the location of the resource on the Internet.

3. Core Meta-Data Concepts

Within the information industry, meta-data has long been used to facilitate access to information. For electronic information exchange, there have been multiple independent initiatives to develop meta-data within various fields of study.

As the amount of information on the Internet grew, it was recognized that a "core" set of meta-data used by all communities was desirable and would improve access to relevant, networked information.

Below is a table summarizing some meta-data specifications reviewed by the KM Standards Working Group:

Name	Origin	No. of available elements	Strengths	Weaknesses
Dublin-Core	W3C (World-Wide Web Consortium)	15	Simple, multi-disciplinary & widely supported	Early in life cycle and still evolving
MARC (Machine Readable Cataloguing)	Library	100 +	Mature	Very complex element set
JPL EIS Library (Docushare)	JPL Administrative	30	Operational system	JPL Project Specific
JPL's PDS	JPL Scientific		Established and operational	JPL Science Data Specific
RDF (Resource Description Framework) XML based	W3C (World-Wide Web Consortium)	NA	Highly extensible and powerful; an integral part of new XML community	Would require the development of a JPL specific core meta-data spec
SOIF (Standards Object Interchange Format)	Internet Working Group	44	Draft Internet Standard	Not widely accepted
GILS	US Govt	25	US Govt backed	Not widely accepted outside Govt
GLU	CDS (French Astronomy Community)	20	Mature and extensible	Astronomy specific

In building infrastructure, it is in JPL's interest to follow the lead of standards developed by an open standards process. It is common to view ourselves and our organizations as unique and reason then that standards don't really apply to our special circumstances. This is a trap. This leads down a path in which local optimizations drive infrastructure architectures and developments. Making decisions that align with broad industry standards results in vendor independence, enables cross-platform support and guarantees long-term data reuse.

4. Core Meta-Data Recommendations

The recommendation from the KM Standards Working Group is that JPL adopt the Dublin Core Element Set from the W3C as outlined in Internet RFC 2413 (referred by name as Qualified Dublin Core or DC 2.0). Its intent is to facilitate discovery of electronic resources. Originally conceived for author-generated description of Web resources, it has been implemented within formal resource description communities within government agencies, museums, libraries, universities, and commercial organizations world-wide (see Section 5, Dublin Core Meta-data Case Studies).

The Dublin Core Workshop Series has gathered experts from the library world, the networking and digital library research communities, and a variety of content specialties in a series of invitational workshops. The building of an interdisciplinary, international consensus around a core element set is the central feature of the Dublin Core. The progress represents the emergent wisdom and collective experience of many stakeholders in the resource description arena.

The characteristics of the Dublin Core that distinguish it as a prominent candidate for description of electronic resources fall into several categories:

Simplicity

The Dublin Core is intended to be usable by non-catalogers as well as resource description specialists. Most of the elements have commonly understood semantics of roughly the complexity of a library catalog card.

Semantic Interoperability

In the Internet Commons, disparate description models interfere with the ability to search across discipline boundaries. Promoting a commonly understood set of descriptors that helps to unify other data content standards increases the possibility of semantic interoperability across disciplines.

World-wide Consensus

Recognition of the international scope of resource discovery on the Web is critical to the development of effective discovery infrastructure. The Dublin Core benefits from active participation and promotion in some 20 countries in North America, Europe, Australia, and Asia.

Extensibility

The Dublin Core provides a streamlined alternative to more elaborate description models such as the full MARC cataloging of the library world. Additionally, it includes sufficient flexibility and extensibility to encode the structure and more elaborate semantics inherent in richer description standards

Meta-data Modularity on the Web

The diversity of meta-data needs on the Web requires an infrastructure that supports the coexistence of complementary, independently maintained meta-data packages. The World Wide Web Consortium (W3C) has begun implementing an architecture for meta-data for the Web. The Resource Description Framework, or RDF, is designed to support the many different meta-data needs of vendors and information providers. Representatives of the Dublin Core effort are actively involved in the development of this architecture, bringing the digital library perspective to bear on this important component of the Web infrastructure.

4.1 The Core Meta-data Element Set

The DC 15 element meta-data set descriptions consist of the following:

1. Title Label: TITLE

The name given to the resource by the CREATOR or PUBLISHER.

2. Author or Creator Label: CREATOR

The person or organization primarily responsible for creating the intellectual content of the resource. For example, authors in the case of written documents, engineers who build performance or design models, illustrators or photographers in the case of visual resources.

3. Subject and Keywords Label: SUBJECT

The topic of the resource. Typically, the subject will be expressed as keywords or phrases that describe the subject or content of the resource. The use of controlled vocabularies and formal classification schemas is encouraged and can be outlined in discipline specific schemas.

4. Description Label: DESCRIPTION

A textual description of the content of the resource, including abstracts in the case of document-like objects or content descriptions in the case of visual resources.

5. Publisher Label: PUBLISHER

The entity responsible for making the resource available in its present form, such as a Division/Section, Project, or Contract -- i.e. who is sponsoring or funding the work.

6. Other Contributor Label: CONTRIBUTOR

A person or organization not specified in a CREATOR element who has made significant intellectual contributions to the resource but whose contribution is secondary to any person or organization specified in a CREATOR element (for example, editor, transcriber, and illustrator).

7. Date Label: DATE

A date associated with the creation or availability of the resource. Such a date is not to be confused with one belonging in the Coverage element, which would be associated with the resource only insofar as the intellectual content is related to that date. Recommended best practice is defined in a profile of ISO 8601 [Date and Time Formats (based on ISO8601), W3C Technical Note, `http://www.w3.org/TR/NOTE-datetime`] that includes (among others) dates of the forms YYYY and YYYY-MM-DD. In this scheme, for example, the date 1994-11-05 corresponds to November 5, 1994.

8. Resource Type Label: TYPE

The category of the resource, such as home page, novel, poem, working paper, technical report, essay, dictionary. For the sake of interoperability, TYPE should be selected from an enumerated list that is under development in the workshop series at the time of publication of this document. See `http://sunsite.berkeley.edu/Metadata/types.html` for current thinking on the application of this element.

9. Format Label: FORMAT

The data format of the resource, used to identify the software and possibly hardware that might be needed to display or operate the resource. For the sake of interoperability, FORMAT should be selected from an enumerated list that is under development in the workshop series at the time of publication of this document.

10. Resource Identifier Label: IDENTIFIER

String or number used to uniquely identify the resource. Examples for networked resources include URLs and URNs (when implemented). Other globally-unique identifiers, such as International Standard Book Numbers (ISBN) or other formal names would also be candidates for this element in the case of off-line resources.

11. Source Label: SOURCE

A string or number used to uniquely identify the work from which this resource was derived, if applicable. For example, a PDF version of a novel might have a SOURCE element containing an ISBN number for the physical book from which the PDF version was derived.

12. Language Label: LANGUAGE

Language(s) of the intellectual content of the resource. Where practical, the content of this field should coincide with RFC 1766.

See: `http://ds.internic.net/rfc/rfc1766.txt`

13. Relation Label: RELATION

The relationship of this resource to other resources. The intent of this element is to provide a means to express relationships among resources that have formal relationships to others, but exist as discrete resources themselves. For example, images in a document, chapters in a book, or items in a collection.

14. Coverage Label: COVERAGE

The spatial and/or temporal characteristics of the resource. Formal specification of COVERAGE is currently under development.

15. Rights Management Label: RIGHTS

A free text explanation pertaining to the terms of access to the resource, or a link to a copyright notice, to a rights-management statement, or to a service that would provide information about terms of access to the resource.

> **Note:** JPL's ITAR Notification could be inserted here as a flag for ITAR restricted information or JPL's formal copyright notice could be inserted for applicable material

There are several important points to remember concerning DC implementations:

1. Not all 15 elements are "required" – to be consistent with the proposed core meta-data specification, you are not required to include every element of the DC set; you simply need to follow the correct syntax and usage for the elements used

2. Elements can repeat – as needed any of the 15 elements can repeat such as is the case with multiple authors (see examples in Section 4.2)

3. DC meta-data can be represented in different formats – HTML, XML/RDF and as a standalone adjunct file. The following section outlines in more detail how the recommended core data spec can be represented in these formats.

It is important to note that the implementation of the core meta-data specification is flexible and tailorable to domain specific applications.

4.2 Examples of DC Element Set in Different Formats

An example of a DC element set embedded in an HTML document is:

```
<!doctype html public "-//w3c//dtd html 4.0 transitional//en">
<html>
<head>
<LINK REL="schema.DC" HREF="http://purl.org/dc">
<META NAME="DC.Title" CONTENT="Using UML to Build Business
Applications">
<META NAME="DC.Creator" CONTENT="Else-Marie Ostling">
<META NAME="DC.Subject" CONTENT="UML; Use Case; Use Case Analysis;
Scenarios; Object Oriented
Software Engineering">
<META NAME="DC.Description" CONTENT="Describes how UML can be applied
to application
development">
<META NAME="DC.Publisher" CONTENT="Distributed Computing Jan/Feb 99
Issue">
<META NAME="DC.Contributor" CONTENT="--">
<META NAME="DC.Date" SCHEME="WTN8601" CONTENT="1999-02-01">
<META NAME="DC.Type" CONTENT="Text">
<META NAME="DC.Format" CONTENT="text/html; 11670 bytes">
<META NAME="DC.Identifier" CONTENT="ARTICLE_Using-UML-to-Build-
Applications.htm">
<META NAME="DC.Source" CONTENT="--">
<META NAME="DC.Language" CONTENT="English">
<META NAME="DC.Relation" CONTENT="--">
<META NAME="DC.Coverage" CONTENT="--">
<META NAME="DC.Rights" CONTENT="Copyright by Computing Magazine, all
rights reserved">
.
.
.
.
```

An example of a DC element set in RDF format is:

```
<rdf:RDF
     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
     xmlns:dc="http://purl.org/dc/elements/1.0/">
      xmlns:dc="">
   <rdf:Description about = ""
     dc:Title = "The Dublin Core Element Set"
     dc:Creator = "The Dublin Core Metadata Inititative"
     dc:Description = "The Dublin Core is a simple metadata element
         set intended to facilitate discovery of electronic resources."
```

```
     dc:Date = "1995-03-01"/>
</rdf:RDF>
```

Many resources do not have the ability to include hyperlinks and/or cannot embed HTML meta-tags or XML RDF tags. These are sometimes referred to as “non-web resources,” e.g., TXT, MOV, AVI, MP3, MIDI, PDF, GIF, JPG, TIFF, PNG, etc. Separate adjunct files can be created that contain structured meta-data following the DC element set.

> **Note:** some of these formats include the capability to embed meta-data and can even include qualified DC-labeled meta-data but the formats are often binary-based making the methods of meta-data extractions unique and sometimes difficult for each format. As a result, adjunct meta-data files are created for each non-web resource adhering to the DC element naming scheme.

There is an Internet draft standard that outlines the structure of a standalone adjunct file called a Summary Object Interchange Format (SOIF) that is described in a 1997 draft RFC. A SOIF is an ASCII-text file that search engines can index. Below is an adjunct meta-data file conforming to the SOIF specification that includes DC meta-data elements:

```
@Dublin-Core-1 { ftp://ds.internic.net/internet-drafts/draft-kunze-dc-
00.txt
TITLE{52}:      Dublin Core Metadata for Simple Resource Description
CREATOR-1{9}:   S. Weibel
CREATOR-2{8}:   J. Kunze
CREATOR-3{9}:   C. Lagoze
PUBLISHER{31}:  Internet Engineering Task Force
CONTRIBUTOR-1{11}:      Nick Arnett
CONTRIBUTOR-2{15}:      Eliot Christian
SUBJECT{44}:    The Dublin Core Set of Elements for Metadata
DESCRIPTION{46}:        Reference description of Dublin Core elements.
CONTRIBUTOR-3{14}:      Martijn Koster
CONTRIBUTOR-4{18}:      Christian Mogensen
CONTRIBUTOR-5{14}:      Timothy Niesen
CONTRIBUTOR-6{11}:      Andrew Wood
CONTRIBUTOR-7{10}:      Mic Bowman
CONTRIBUTOR-8{11}:      Dan Connoly
CONTRIBUTOR-9{15}:      Michael Mauldin
CONTRIBUTOR-10{12}:     Wick Nichols
DATE{16}:       February 9, 1997
TYPE{14}:       Internet draft
FORMAT{4}:      Text
IDENTIFIER:{21} draft-kunze-dc-00.txt
SOURCE{41}:     http://purl.oclc.org/metadata/dublin_core
LANGUAGE{3}:    eng
RELATION{24}:   Draft Reference Standard
COVERAGE{22}:   Expires August 8, 1997
RIGHTS{58}:     Unlimited Distribution; readers must not cite as
standard.}
```

or a more familiar document:

```
@Dublin-Core-1 { ftp://ds.internic.net/internet-drafts/draft-kunze-dc-
00.txt
TITLE:            The Declaration of Independance
CREATOR-1:        Thomas Jefferson
PUBLISHER:        The Second Continental Congress
```

```
CONTRIBUTOR-1:    Benjamin Franklin
CONTRIBUTOR-2:    John Adams
CONTRIBUTOR-3:    Roger Sherman
CONTRIBUTOR-4:    Robert R. Livingston
SUBJECT:          The inalienable rights of the citizens of the
colonies
DESCRIPTION:      A treatise outlining the rights of the colonies.
DATE.COMPLETED    1776-07-04
DATE.SIGNED:      1776-08-02
TYPE:             Political treatise
FORMAT:           Hardcopy; parchment
SOURCE:           National Archives, Washington, DC
LANGUAGE:         eng
RELATION:         Based on inalienable God-given rights
COVERAGE:         The United States of America
RIGHTS:           Unlimited Distribution to a
```

Below is an example of a variation of a SOIF adjunct meta-data file for a piece of software:

```
%FAST File Search 1.0
Content-Type: text/plain; charset=”iso-8859-1”
Content-Transfer-Encoding: 8BIT
DC.Title: Learning the Planets
DC.Creator: Bob Rock
DC.Date: 1995-12-18
DC.Type: Software
DC.Format: Applet
DC.Identifier: ftp://ftp.sample.org/pub/educational-
software/ARCHIVE_planets12.zip
DC.Language:
DC.Rights: shareware
FAST.Price: 200; scheme=”NOK”
DC.Subject: Java; Educational; Astronomy, Planetary Astronomy
DC.Description: Animated software that helps you learn the planets of
the solar system.
Great graphics. Requires sound support. This is a educational game for
kids aged 6 to 9.
FAST.Proglanguage: Java
FAST.Sourcecode: YES
FAST.Environment: Internet
```

4.3 Implementation Issues

Implementing a Core Meta-data Specification does not require radical changes in work practices or information tool-sets. It requires changes in the way information tools are configured and supported; it requires minor and subtle changes in the way users currently use information tools and make information available; it requires some institutional infrastructures to be adapted.

Specific follow-on topics that need to be pursued following the adoption of the Core Meta-data Specification are:

Tuning the Core Meta-data Spec for JPL

Using the core meta-data specification, as outlined in this paper, is relatively straightforward. The problem becomes more complex as discipline and application-specific extensions are required. There are two methods used for extending meta-data structures for specific applications: (1) qualifiers and (2) namespace extensions.

Qualifiers refine elements within the core meta-data namespace. With only 15 elements to the Core Meta-data specification, there is sometimes a need to provide multiple values for a given element and these values need to be differentiated from one another. For example, in the date element, there may be a need to identify different types of dates; e.g., date created, date last modified, date no longer valid.

Using a proposed qualifier scheme (R. Guenther/1997), different dates could be represented in the following ways in HTML:

```
<META NAME="DC.Date.Created" SCHEME="WTN8601" CONTENT="1999-02-01">
<META NAME="DC.Date.Modified" SCHEME="WTN8601" CONTENT="1999-05-07">
<META NAME="DC.Date.ValidTo" SCHEME="WTN8601" CONTENT="1999-12-31">
```

Note: qualifier schemes should be managed and maintained in the same manner as any part of a meta-data namespace.

Additional application- or discipline-specific schemas (namespaces) can be created using the W3C's RDF syntax and Dublin Core extensions as outlined in Internet RFC 2413 entitled Dublin Core Metadata for Resource Discovery . Additional discipline-specific namespaces would include meta-data elements unique to the discipline. For example, engineering data namespace could include reference designation, version, revision, CAGE No., etc. while a science data namespace could include instrument ID, encounter, sequence, range, etc. Additional namespaces are identified via the same syntax used in identifying the core meta-data source, e.g.:

```
<LINK REL="schema.DC" HREF="http://purl.org/dc">
```

or if JPL needs to refine the Dublin Core specification:

```
<LINK REL="schema.DC" HREF="http://data-dictionary.jpl.nasa.gov/core-meta-
data">
```

Discipline specific namespaces could be identified as for the namespace assocated with JPL's Product Data Management System:

```
<LINK REL="schema.DC" HREF="http://data-dictionary.jpl.nasa.gov/pdms">
```

or for the namespace associated with JPL's Planetary Data System:

```
<LINK REL="schema.DC" HREF="http://data-dictionary.jpl.nasa.gov/pds">
```

Note: the Dublin Core Meta-data element set is meta-data specification in contrast to the W3C's RDF (Resource Description Framework), which, as its name implies, is a syntax framework for creating meta-data namespaces. In short the DC element set is the starting point and RDF contains the methods for extending from that starting point.

Establishing an Online, Real-time Data Dictionary Service

Once a core meta-data specification is adopted and as discipline-specific namespaces are developed, an on-line data dictionary service needs to be created to manage these resources. A similar resource already exists for managing the namespace for computer names – it is known as the DNS (Domain Name Service). There are prototype systems in development that address these issues. One, developed under ARPA funding, is call the Handle System. `<URL: http://www.handle.net/>`

Configuring Search Engines to Take Advantage of Tagged and Embedded Meta-data

Making on-line resources easier to find and access is one of the primary reasons for adopting a core meta-data specification but for this capability to be realized, search engines will have to be configured to take advantage of these structures. Both Harvest, a public domain search engine, and Verity, a commercial product, can build search indexes based on embedded meta-data.

Need for Continuous Review and Evaluation of Standards

As information technology continues to evolve at an ever increasing pace, there is a critical need for ongoing review and evaluation of relevant standards. This internal process should match the open consensus based processes employed by organizations such as the Internet IETF Working Groups and the World Wide Web Consortium. The process should include a maturity gating mechanism in which standards are reviewed, evaluated, recommended, tested and then implemented. This approach balances community participation and timely delivery of practical solutions.

5. Multi-disciplinary Meta-data Case Studies

The following are summaries of a few Dublin Core meta-data implementations that demonstrate the wide applicability and support of the recommended core meta-data specification:

Digital Library Catalog – University of California, Berkley

This project, built by UC Berkley and sponsored by Sun MicroSystems, includes books, essays, speeches, and other textual material in HTML, technical reports (in various formats), photographs, engravings and other visual materials, and video and sound clips using DC meta-data structures.

`<URL: http://sunsite.berkley.edu.Catalog>`

Physics Oriented Meta-data Repository – University of Oldenburg

MetaPhys is a Meta Search Engine which serves as a unified query interface to the databases of various publishers with relevance in physics. Built and maintained by the University of Oldenburg in Germany and sponsored by the European Physical Society, MetaPhys includes a simple meta-data authoring tool and meta-data search engine. The goal of this project is to build up a distributed electronic information system in physics.

`<URL:http://www.physik.uni-oldenburg.de>`

Australian Geodynamics Cooperative Research Centre - La Trobe University

The AGCRC, a collaboration between two public research organizations and two universities, is using the WWW as a primary delivery system for the results of its research. The results are composed of materials in a variety of formats that are presented as different resource types. The project is using two different metadata systems for text and numeric data and a link has been made between them. The goal of the project is to ensure that greater research results are produced through effective cooperation between research institutions.

`<URL: http://www.agcrc.csiro.au>`

Subject Area Information Guide for Mathematics, Earth Sciences and History - University of Hamburg

Meta-data is generated for the listing and evaluation of information related to Mathematics, Earth Sciences, or Anglo-American History and Literature. Sources include Internet servers, CD-ROMs and reference books.

`<URL: http://www.sub.uni-goettingen.de/ssgfi/index.html>`

Electronic Library Image Service - De Montfort University

The ELISE service, managed by the International Institute for Electronic Library Research (IIELR) at De Montfort University (UL), operates on a client/server model, making use of z39.50 and Dublin Core. In the ELISE II prototype, the catalogue data supplied by participating institutions is mapped to DC and includes text, image, audio, and video information.

`<URL: http//severn.dmu.ac.uk/elise/>`

National Spatial Data Infrastructure - Federal Geographic Data Committee

The Federal Geographic Data Committee (FGDC) coordinates the development of the National Spatial Data Infrastructure (NSDI). The NSDI encompasses policies, standards, and procedures for organizations to cooperatively produce and share geographic data. The FGDC Meta-data Working Group actively promotes the awareness among FGDC member agencies of the meta-data dimension to geospatial data; facilitates the evolution and revision of the Content Standard for Digital Geospatial Meta-data (CSDGM); and establishes a mechanism for the coordination, development, use, sharing, and dissemination of geospatial meta-data among FGDC member agencies.

`<URL:http://www.fgdc.gov/index.html>`

Medical Meta-data Project - Duke University and IETF

Duke University, the American Medical Informatics Association Internet Working Group, and the National Cancer Institute are working together to provide a test set of the National Cancer Institute Cancer Genetics database and standardized patient record database.

`<URL: http://www.mcis.duke.edu/standards/guide.htm >`

Other projects using Dublin Core Meta-data specification are described at

`<URL: http://purl.oclc.org/dc/projects/index.htm>`

6. Glossary

Attribute/value pair - a model commonly used to define the semantics of meta-data; e.g. in e-mail and HTTP messages, meta-data information is passed in which To, From, Date, Subject, Organization, etc. and are are attributes with associated information as the value.

Binding - the mechanism that makes the association between meta-data and the resource it describes

CDF - Channel Definition Format. An XML application, primarily for describing ‘push’ channels. May also be used to describe arbitrary groupings of resources.

CORBA - Common Object Request Broker Architecture.

DC - See: Dublin Core.

DC-dot - A program for creating Dublin Core meta-data which can be pasted into the headers of Web pages.

DN - Distinguished Name. An X.500 or LDAP object name.

DOI - Digital Object Identifier. An identification system based on Handle System technology, it identifies intellectual property in the digital environment. Developed by the International DOI Foundation on behalf of the publishing industry, its goals are to provide a framework for managing intellectual content, link customers with publishers, facilitate electronic commerce, and enable automated copyright management.

DTD - Document Type Definition. An application program defining document types in an SGML context.

Dublin Core - A meta-data format defined on the basis of international consensus which outlines a minimal information resource description, generally for use in a WWW environment.

EAD - Encoding Archival Description. An SGML-based meta-data format developed for the description of archives.

EDI - Electronic Data Interchange. The exchange of structured data messages to enable automated transactions between application systems.

GILS - Government Information Locator Service. Meta-data format created by the US Federal Government in order to provide a means of locating information generated by government agencies.

Handle System -   A distributed system developed by CNRI which resolves identifiers into the information necessary to locate and access resources.

Harvest - a search engine system developed by an IETF working group providing an architecture for gathering (agent), indexing and accessing Internet information. Uses SOIF.

HTML - Hypertext Mark-up Language. The standard language used for creating Web documents.

HTTP - HyperText Transfer Protocol. The client-server protocol used for the exchange of HTML.

IETF - Internet Engineering Task Force.

ISBN - International Standard Book Number.

ISSN - International Standard Serial Number.

ITAR – International Trade in Arms & Restrictions.

LDAP (RFC 1777) **-** Lightweight Directory Access Protocol. Internet standard for directory services.

MARC (ISO 2709) **-** Machine Readable Cataloguing. A family of formats based on ISO 2709 for the exchange of bibliographic and other related information in machine readable form. For example, USMARC, UKMARC and UNIMARC.

Meta-data - data which provides information about a resource; information that describes information; machine understandable information about web resources or other things [T. Berners-Lee].

Namespace – (general) a name or group of names that are defined according to some naming convention; (Meta-data) a mechanism for providing human and/or machine-readable context for a resource description element. In HTML, namespace(s) are identified via a meta-tag entry:

```
<LINK rel="schema.DC"href="http://www.w3.org/dc/elements/1.0">
```

In RDF, namespaces are identified via a header entry:

```
xmlns:dc="http://www.w3.org/dc/elements/1.0"
```

NIIIP - National Industrial Information Infrastructure Protocols - a team of organizations that has entered into a cooperative development agreement with the U.S. Government (ARPA) to develop open industry software protocols that will make it possible for manufacturers and their suppliers to effectively interoperate as if they were part of the same enterprise.

PICS - Platform for Internet Content Selection - an Internet content filtering infrastructure.

Property - a specific aspect, characteristic, attribute or relation used to describe a resource. Each property has a specific meaning, defines its permitted values, the types of resources it can describe, and its relationship with other properties (W3C).

Push technology - Web technologies based on pushing information (server-side) to end users (client-side); as opposed to more conventional Web-based client-side requests of servers.

Resource - anything being described by meta-data; may or may not be in electronic format, e.g. a book is a resource described by a card in a card catalogue system.

RDF - Resource Description Framework. RDF is currently under development within the W3C and provides a framework for meta-data in different application areas, e.g. resource discovery, content ratings and intellectual property.

RFC – Request for Comment. An Internet document used to outline proposed and approved Internet standards.

SGML (ISO 8879) **-** Standard Generalized Mark-up Language. An international standard for the description of marked-up electronic text.

SiRPAC - RDF-XML parser developed by W3C.

SOIF - Summary Object Interchange Format. A meta-data format developed for use with the Harvest architecture; an Internet Draft Standard, January, 1997, developed by Internet Anonymous FTP Archives

IETF Working Group (IAFA).

Statements - in meta-data modeling, statements consist of a resource together with a named property plus the value of that property; these components are called, respectively, the subject, predicate and object. (W3C)

Unicode - A standard for international character encoding.

URC - Uniform Resource Characteristics.

URI - Uniform Resource Identifier. The super-set of URNs, URLs and URCs.

URL - Uniform Resource Locator. The standard way to give the address of a source of information on the WWW. For example:

```
http://www.ukoln.ac.uk/metadata/publications.html
```

URN - Uniform Resource Name. Persistent identification for Web resources.

Warwick Framework - An architecture for the exchange of distinct meta-data packages involving the aggregation of meta-data packages into containers.

Web robot - A software robot which trawls the WWW, generating all-encompassing Web indexes. Also known as Web crawlers or Web spiders.

WebDAV - Web Distributed Authoring and Versioning. An IETF Working Group.

World Wide Web Consortium - Organization currently responsible for the development of Web protocols.

WWW -World Wide Web.

W3C- See World Wide Web Consortium

XML - Extensible Markup Language. A lightweight version of SGML designed for use on the Internet.

Z39.50 - NISO standard for an applications layer protocol for information retrieval which is specifically designed to aid retrieval from distributed servers.

7. References

"Resource Description Framework (RDF) Model and Syntax Specification," World Wide Web Consortium, February 22, 1999,

Dublin Core Meta-data Element Set: Reference Description, DC Meta-data Initiative (a W3C Working Group), October 2, 1997, `<URL:http://oclc.org/dc/about/element_set.htm>`

"CIP Index Object Format for SOIF Objects" - an INTERNET-DRAFT RFC, April, 1998, Edward Hardie/NASA NIC, M. Bowman/Transarc, D. Hardy/Netscape, M. Schwartz @Home, D.Wessels/NLANR

S. Weibel, J. Kunze, C. Lagoze, Dublin Core Meta-data for Simple Resource Description, February 1997, <URL:ftp://ds.internic.net/internet-drafts/draft-kunze-dc-00.txt>

E. Miller, Dublin Core Element Set Crosswalk, January 1997, <URL:http://www.oclc.org:5046/~emiller/DC/crosswalk.html>

The Harvest Information Discovery and Access System, <URL:http://harvest.transarc.com/>

Berners-Lee, Meta-data Architecture: Documents, Meta-data, and Links, January 1997, <URL: http://www.w3.org/DesignIssues/Metadata.html>

G. Hannemyr, Discovering Resources on the Internet, May 17, 1997.

J. Hughes and A. Farny, The Management of Meta-data within the Planetary Data System, <URL: http://www.llnl.gov/liv_comp/metadata/papers/pos-papers-may-1994/StevenHughes>

8. Contributions

Knowledge Management Standards Working Group

Ann Bernath / 389

Kris Blom / 390

Robin Dumas / 389

Elizabeth Duxbury / 388

Don Fuhrman / 351

Jennifer Momjian / 643

Julie Reiz / 311

Chris Green / 389

Susan Roberts / 363

Chuck Shinbrot / 501

Sugi Sorensen / 389

Editor

Ed Sewall

Other Contributors

Ruth Bergman / 311